Evaluation Dataset for Japanese Medical Text Simplification

Koki Horiguchi† **Tomoyuki Kajiwara**† **Yuki Arase**‡ **Takashi Ninomiya**†
† Graduate School of Science and Engineering, Ehime University, Japan
‡ Graduate School of Information Science and Technology, Osaka University, Japan
{horiguchi@ai., kajiwara@, ninomiya@}cs.ehime-u.ac.jp
arase@ist.osaka-u.ac.jp

Abstract

We create a parallel corpus for medical text simplification in Japanese, which simplifies medical terms into expressions that patients can understand without effort. While text simplification in the medial domain is strongly desired by society, it is less explored in Japanese because of the lack of language resources. In this study, we build a parallel corpus for Japanese text simplification evaluation in the medical domain using patients' weblogs. This corpus consists of $1,425$ pairs of complex and simple sentences with or without medical terms. To tackle medical text simplification without a training corpus of the corresponding domain, we repurpose a Japanese text simplification model of other domains. Furthermore, we propose a lexically constrained reranking method that allows to avoid technical terms to be output. Experimental results show that our method contributes to achieving higher simplification performance in the medical domain.

1 Introduction

Medical texts contain a lot of technical terms which are often difficult to understand for laypeople (Cheng and Dunn, 2015). For better communication between medical practitioners and patients, medical text simplification has been actively studied in English so that the patients understand their diseases and symptoms and the courses of treatments in detail (Cao et al., 2020; Sakakini et al., 2020; Guo et al., 2021; Devaraj et al., 2021; Luo et al., 2022). However, such studies have not been well explored in Japanese because of the lack of a parallel corpus of this domain.

In the case of English, a typical approach is to employ pre-trained models in medical domain (Lee et al., 2019; Alsentzer et al., 2019) or models specialized for text simplification (Sun and Wan, 2022; Sun et al., 2023) such as SimpleBERT. However, there have not been any off-the-shelf pre-trained models of these kinds in Japanese.

As a step-forward to Japanese medical text simplification, we create and release a parallel corpus for evaluation, named JASMINE.[1] In addition, we tackle this problem without any domain-specific training corpus by developing a Japanese version of SimpleBART,[2] a pre-trained model specialized for text simplification. Furthermore, we propose a reranking method that avoids technical terms and employ it to the text simplification model. Experimental results on our JASMINE corpus confirm efficacy of out methods.

2 Related Work

2.1 Parallel Corpus for Text Simplification

In text simplification, a monolingual parallel corpus consisting of pairs of complex and simple sentences is essential to train and evaluate seq2seq models. In English, a large-scale parallel corpus has been built by automatic sentence alignment from article pairs with different target readers from Wikipedia (Jiang et al., 2020) and news (Xu et al., 2015). In Japanese, SNOW[3,4] (Maruyama and Yamamoto, 2018; Katsuta and Yamamoto, 2018), which is manually simplified sentences from textbooks, and JADES[5] (Hayakawa et al., 2022), which is similarly constructed from news, are publicly available.

Medical text simplification is the task of paraphrasing texts written by medical practitionars into expressions easy to understand for patients by avoiding technical terms and phrases. Although medical text simplification has been actively studied in English (Cao et al., 2020; Devaraj et al.,

[1] JASMINE: **JA**panese text **S**implification dataset in the **M**edical doma**IN** for **E**valuation. https://github.com/OnizukaLab/JASMINE
[2] https://github.com/EhimeNLP/JapaneseSimpleBART
[3] https://www.jnlp.org/GengoHouse/snow/t15
[4] https://www.jnlp.org/GengoHouse/snow/t23
[5] https://github.com/naist-nlp/jades

Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 4: Student Research Workshop), pages 219–225
June 18, 2024 ©2024 Association for Computational Linguistics

Complex	体調不良は、顔面浮腫と酔っ払った感覚が強くなっているためだろう。
(I suffer from ill health probably due to facial edema and a drunken feeling getting worse.)	
Simple	体調悪いのは、顔面のむくみと酔っ払った感覚が強くなっているからだと思われる。
(I am feeling under the weather, likely because facial swelling and drunken feeling are getting worse.)	

Table 1: An example of our parallel corpus. Technical terms are in red, and expressions that simplify them for patients are in blue. Underlined parts are examples of style transfer.

2021; Luo et al., 2022), less efforts have been made for Japanese because of lack of parallel corpus of this domain.

2.2 Pre-training for Text Simplification

Transfer learning of a pre-trained model, where the model trained by a large-scale raw corpus is fine-tuned using a corpus of target task, has been successful in various natural language processing tasks. For text-to-text generation tasks, such as text simplification, BART (Lewis et al., 2020), a pre-trained Transformer (Vaswani et al., 2017) trained via denoising autoencoding, is widely used and demonstrated its effectiveness (Martin et al., 2022; Hatagaki et al., 2022; Zetsu et al., 2022).

Recent studies have shown that task-specific pre-training is more effective. For example, the effectiveness of pre-training to mask and reconstruct sentences for summarization (Zhang et al., 2020) and pre-training to reconstruct round-trip translations for paraphrasing (Kajiwara et al., 2020) have been reported. For text simplification, continued pre-training of BART with masked language modeling with focus on simple words, released as SimpleBART (Sun et al., 2023), improves the simplification quality. Despite its success in English, Japanese version of SimpleBART has not yet been developed.

3 JASMINE Corpus

To enable the evaluation of medical text simplification in Japanese, we create a parallel corpus, named JASMINE[1], consisting of sentences with and without medical terms. We employ patients' weblog articles written primarily for daily records of symptoms and treatments and communication with other patients[6] as medical texts for non-professional use (henceforth, it will be referred to simply as a blog). We construct a parallel corpus consisting of pairs of complex and simple sentences with or without technical terms by manually paraphrasing the (simple) blog sentences to use medical terms in MedDRA.[7] As shown in Table 1, style transfer from blog-specific informal expressions to formal ones is also performed during paraphrasing.

3.1 Manual Paraphrasing

We hired two annotators with over 20 years of annotation experience in Japanese natural language processing. One annotator with experience in annotating medical domains (not a medical professional) paraphrased simple blog sentences, and the other annotator checked and corrected them. The annotators made sure to replace disease names and symptoms written by lay patients with corresponding medical terms in the MedDRA dictionary. Our corpus is built from $2,009$ sentence pairs from blog articles. It is notable that these sentences were extracted from approximately $17,000$ sentences in $1,000$ blog posts, which confirms the scarcity of available resources for Japanese medical simplification.

3.2 Correction and Filtering

The $2,009$ pairs obtained in the previous section included inappropriate examples such as sentence fragments and pairs that are too much context dependent due to article-level paraphrasing by annotators. We manually checked all pairs to exclude these samples and made further formatting corrections where applicable, such as complementing missing punctuation marks at the end of sentences and removing emoticons. Finally, we obtained $1,425$ sentence pairs of Japanese parallel corpus for medical text simplification as JASMINE.

Our JASMINE corpus is not the scale for training or fine-tuning a model. However, it preserves a sufficient size for evaluating medical text simplification models compared to the sizes of the existing evaluation sets for general text simplification: 359 sentence pairs in TurkCorpus (Xu et al., 2016) and ASSET (Alva-Manchego et al., 2020) (in English)

[6] https://www.tobyo.jp/

[7] We used the online medical term dictionary: http://sociocom.jp/~data/2018-manbyo/

and 100 sentence pairs for SNOW (Maruyama and Yamamoto, 2018; Katsuta and Yamamoto, 2018) (in Japanese).

4 Japanese Simplification Model

For achieving medical text simplification in Japanese, we develop a pre-trained model specific to the text simplification task (Section 4.1) and propose a reranking method for avoiding technical terms (Section 4.2).

4.1 Japanese SimpleBART

Following previous work in English (Sun et al., 2023), we develop Japanese SimpleBART[2] with an enriched ability to generate simple words. Specifically, SimpleBART conducts continued pre-training using the text simplification parallel corpus on BART. Note that these SimpleBART models acquire text simplification abilities after further fine-tuning with the text simplification parallel corpus. The rest of this section describes continual pre-training methods using simple sentences and complex sentences, respectively.

Masking for Simple Sentences For continual training of SimpleBART, simple words are used as masking targets. For this purpose, previous work in English (Sun et al., 2023) has used a complex word identification model (Pan et al., 2021), whereas a large-scale word difficulty lexicon[8] (Nishihara and Kajiwara, 2020) is available for Japanese. The lexicon consists of $40,605$ Japanese words, each of which is assigned three levels of difficulty (easy, medium, and difficult). In this study, both easy and medium words are defined as simple words and are masked.

Following the previous studies (Lewis et al., 2020; Sun et al., 2023), we mask 15% of the words among those to be masked. However, since it is unlikely that the mask is applied to sentences containing many difficult words or words not registered in the lexicon, the percentage of masked target words t is considered for each sentence and adjusted so that 15% of the words are actually masked in each sentence. In addition, to mask simpler words more frequently, we multiply the mask probability by the weight of $0 \leq \theta \leq 1$. In this study, $\theta = 1$ for easy words, $\theta = 0.75$ for medium words, and $\theta = 0$ for other words. Finally, the masking probability m is as follows.

$$m = \min(\frac{0.15\theta}{t},\ 1.0) \quad (1)$$

Masking for Complex Sentences In the continual training of SimpleBART, we aim to reduce the generation of complex words as well as promote the generation of simple words. For this purpose, Sun et al. (2023) mask complex words in sentences and reconstruct their simple synonyms using the lexical simplification lexicon, SimplePPDB++ (Maddela and Xu, 2018). Similarly, we use the Japanese version of the lexical simplification lexicon[8] (Nishihara and Kajiwara, 2020). While the lexicon consists of $42,642$ word pairs, only $18,810$ word pairs with a cosine similarity of word embeddings[9] greater than 0.25 are used as reliable simplification. If there are multiple simplification candidates, the candidate with the highest word-filling probability of BERT[10] (Devlin et al., 2019) is selected.

Following previous studies (Lewis et al., 2020; Sun et al., 2023), we mask 15% of the words. Since the mask target is limited to words registered in the lexicon, the mask may be less applicable to some sentences. Therefore, as in the previous section, we dynamically weight the mask probabilities for each sentence using Equation (1). Here, $\theta = 1$ for words registered in the lexical simplification lexicon and $\theta = 0$ for other words.

4.2 Lexically Constrained Reranking

To generate paraphrases that do not include technical terms, we propose a reranking method that selects simplifications that do not contain given words among multiple candidate sentences. Our method first generates multiple candidate sentences using a trained text simplification model with any decoding algorithm, such as beam search or Top-p sampling (Holtzman et al., 2020). These candidate sentences are then checked in the ascending order of their generation ranks, and the first simplification that does not contain any given words is output. However, if the given words are included in all candidate sentences, the first candidate sentence is output. Although our experiment requires that all medical terms in the MedDRA dictionary be avoided, exploration of better lexical constraints remains a future work.

[8] https://github.com/Nishihara-Daiki/lsj

[9] https://cl.asahi.com/api_data/wordembedding.html

[10] https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking

	Train	Valid	Test
SNOW	$82,300$	$1,000$	100
JADES	-	$2,959$	948
JASMINE	-	425	$1,000$

Table 2: Number of sentence pairs for each corpus.

5 Experiments

To evaluate the performance of the medical text simplification, we experiment with the JASMINE that we constructed. We also experiment with SNOW (Maruyama and Yamamoto, 2018; Katsuta and Yamamoto, 2018) and JADES (Hayakawa et al., 2022), which are existing parallel corpora for Japanese text simplification, to evaluate simplification performance in other domains. In JASMINE, we also evaluate the effectiveness of lexically constrained reranking with the MedDRA dictionary.[7] Table 2 shows the corpus sizes.

5.1 Experimental Setup

Model We compare BART[11] pre-trained on the Japanese Wikipedia with the following three models that further apply continual pre-training.

- BART-CP: continual pre-training on SNOW with general masked language modeling.
- SimpleBART: continual pre-training on SNOW with masked language modeling that focuses on the simple words (Section 4.1).
- SimpleBART-CP: continual pre-training on SNOW with general masked language modeling, and then further continual pre-training with masked language modeling focused on simple words.

These pre-trained models were further fine-tuned on SNOW to develop text simplification models.

Continual Pre-training An continual 10 epochs of pre-training was conducted on SNOW. Pre-processing was performed following BART,[11] with word segmentation by Juman++[12] (Tolmachev et al., 2018) and subword segmentation by SentencePiece[13] (Kudo and Richardson, 2018). The batch size was set to 64, the Dropout rate to 0.1, and the optimization method was AdamW (Loshchilov and Hutter, 2019). Polynomial decay was used for learning rate scheduling, with a maximum learning rate of 5×10^{-5} with a warmup step of $5,000$.

Fine-tuning SNOW was also used for fine-tuning. We did the same pre-processing as the continual pre-training. The batch size was set to 64, the dropout rate to 0.3, and AdamW was used for optimization. Polynomial decay was used for learning rate scheduling, with a maximum learning rate of 3×10^{-5} and a warmup step of $2,500$. Fine-tuning was stopped early if no improvement on cross-entropy loss measured using the validation set is observed for 5 epochs.

Inference For evaluation, we generated simple sentences using beam search with a beam size of 5. In JASMINE, we also experimented with generating simple sentences using our lexically constrained reranking. During lexically constrained reranking, we generated candidates for $n = 100$ sentences each by beam search with a beam size of 100 and Top-p sampling with $p = 0.95$.

Evaluation Metric Simplification performance was automatically evaluated using SARI (Xu et al., 2016) with EASSE toolkit[14] (Alva-Manchego et al., 2019). Although SARI was a metric originally proposed for English text simplification, it has also been used for Japanese text simplification (Hatagaki et al., 2022; Hayakawa et al., 2022).

5.2 Results

Experimental results are shown in Table 3. We conducted five experiments with different random seeds and reported the average scores. Bootstrap method was used for statistical significance tests.

The result that SimpleBART consistently outperforms BART confirms the effectiveness of pre-training specialized for Japanese text simplification. While BART-CP also consistently outperforms BART, SimpleBART achieved higher simplification performance in all settings except SNOW. Except in the medical domain, the highest performance was achieved by SimpleBART-CP with continual training in both general masked language modeling and those focused on simple words. These results demonstrate the effectiveness of continual training focused on simple words.

[11] https://huggingface.co/ku-nlp/bart-base-japanese
[12] https://github.com/ku-nlp/jumanpp
[13] https://github.com/google/sentencepiece
[14] https://github.com/feralvam/easse

	SNOW	JADES	JASMINE		
Decoding	Beam	Beam	Beam	Beam	Top-p
Reranking				✓	✓
BART	63.70	36.69	32.88	36.66	35.72
BART-CP	64.27*	37.55*	34.04*	36.80*	36.41*
SimpleBART	64.06*	37.57*	**34.72***	**37.37***	**36.48***
SimpleBART-CP	**64.34***	**39.09***	34.43*	36.80*	36.17*

Table 3: Results of SARI scors (* indicates statistically significant difference from BART ($p < 0.05$))

Input	度重なる**腸閉塞**と**腸管拡張**があったため、結局腹部には効いていないとの判断になった。 (Because of the repeated **intestinal obstruction** and **intestinal enlargement**, (the treatment) has been determined ineffective to my abdomen.)
BART	度重なる腸の閉塞と腸管の拡張があったため、結局腹には効いていないとの判断になった。 (Because of the repeated obstructions in intestine and enlargement of intestinal tract, ...)
SimpleBART	何度も腸の閉塞と腹の臓器の拡張があったため、結局腹部には効果がないとの判断になった。 (Because of the repeated obstructions in intestine and enlargement of the abdominal organs, ...)
+ Reranking	何度も腸の障害があったため、結局腹には効果がないとの判断になった。 (Because of the repeated intestinal disorders, ...)

Table 4: Examples of simplification outputs with "腸閉塞 (intestinal obstruction)" and "腸管拡張 (intestinal enlargement)" as lexical constraints. Successful paraphrases are in blue and those that failed are in red.

In the evaluation on JASMINE, lexically constrained reranking significantly improves the simplification performance. As for decoding methods, beam search consistently outperforms Top-p sampling. Table 5 shows the percentage of output sentences that do not include technical terms when the number of candidate sentences n is varied from 10 to 200 during lexically constrained reranking. We find that lexically constrained reranking can significantly improve the number of output sentences without technical terms, as well as the SARI score of the simplification performance. The greater the number of candidates for reranking, the greater the percentage of sentences without technical terms, but even reranking only 10 sentences has a significant impact.

Table 4 shows an example of our Japanese medical text simplification. For SimpleBART model, we show both output sentences with or without our lexically constrained reranking on top of beam search. This example contains two technical terms "腸閉塞 (intestinal obstruction)" and "腸管拡張 (intestinal enlargement)". SimpleBART with lexically constrained reranking paraphrases them into a simple phrase of "腸の障害 (intestinal disorders)" which is not a technical term. Without lexically constrained reranking, the latter technical term of "腸管拡張 (intestinal enlargement" is paraphrased into "腹の臓器の拡張 (enlargement of the abdominal organs)" but the former technical term remains.

	$n = 1$	$n = 10$	$n = 50$	$n = 100$	$n = 200$
Beam	77.4	91.8	94.9	96.2	97.2
Top-p	80.8	85.0	86.9	87.4	88.2

Table 5: Percentage of output sentences without technical terms when varying the number of candidates n during lexically constrained reranking in SimpleBART. (The beam size was the same as n, but for $n = 1$, the beam size was set to 5.)

6 Conclusion

This study released a set of language resources for medical text simplification in Japanese: an evaluation corpus of JASMINE[1] and a pre-trained model of Japanese SimpleBART.[2] Experimental results in three domains, including medical, show that our Japanese SimpleBART consistently achieves high performance, and reveal the effectiveness of pre-training specialized for text simplification. In the medical domain, our lexically constrained reranking to avoid technical terms further improved the simplification performance.

Our future work includes the construction of a large-scale training parallel corpus for medical text simplification in Japanese. We also plan to examine more sophisticated lexical constraints, such as allowing common technical terms to be output.

Limitations

Since this study is on sentence simplification, our text simplification models are not able to account for context beyond the sentence in our experiments. However, note that the annotations in our dataset do not have such limitations and that the annotators read the entire document. None of our annotators are medical professionals, although one of them has experience in text annotation in the medical domain.

Acknowledgments

This work was supported by MHLW AC Program Grant Number JPMW22AC5001 and Cross-ministerial Strategic Innovation Promotion Program (SIP) on "Integrated Health Care System" Grant Number JPJ012425.

References

Emily Alsentzer, John Murphy, William Boag, Wei-Hung Weng, Di Jindi, Tristan Naumann, and Matthew McDermott. 2019. Publicly Available Clinical BERT Embeddings. In *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, pages 72–78.

Fernando Alva-Manchego, Louis Martin, Antoine Bordes, Carolina Scarton, Benoît Sagot, and Lucia Specia. 2020. ASSET: A Dataset for Tuning and Evaluation of Sentence Simplification Models with Multiple Rewriting Transformations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4668–4679.

Fernando Alva-Manchego, Louis Martin, Carolina Scarton, and Lucia Specia. 2019. EASSE: Easier Automatic Sentence Simplification Evaluation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing: System Demonstrations*, pages 49–54.

Yixin Cao, Ruihao Shui, Liangming Pan, Min-Yen Kan, Zhiyuan Liu, and Tat-Seng Chua. 2020. Expertise Style Transfer: A New Task Towards Better Communication between Experts and Laymen. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1061–1071.

Christina Cheng and Matthew Dunn. 2015. Health Literacy and the Internet: A Study on the Readability of Australian Online Health Information. *Australian and New Zealand Journal of Public Health*, 39(4):309–314.

Ashwin Devaraj, Iain Marshall, Byron Wallace, and Junyi Jessy Li. 2021. Paragraph-level Simplification of Medical Texts. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4972–4984.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186.

Yue Guo, Wei Qiu, Yizhong Wang, and Trevor Cohen. 2021. Automated Lay Language Summarization of Biomedical Scientific Reviews. In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence*, pages 160–168.

Koki Hatagaki, Tomoyuki Kajiwara, and Takashi Ninomiya. 2022. Parallel Corpus Filtering for Japanese Text Simplification. In *Proceedings of the 2022 Workshop on Text Simplification, Accessibility, and Readability*, pages 12–18.

Akio Hayakawa, Tomoyuki Kajiwara, Hiroki Ouchi, and Taro Watanabe. 2022. JADES: New Text Simplification Dataset in Japanese Targeted at Non-Native Speakers. In *Proceedings of the 2022 Workshop on Text Simplification, Accessibility, and Readability*, pages 179–187.

Ari Holtzman, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. 2020. The Curious Case of Neural Text Degeneration. In *Proceedings of the 2020 International Conference on Learning Representations*.

Chao Jiang, Mounica Maddela, Wuwei Lan, Yang Zhong, and Wei Xu. 2020. Neural CRF Model for Sentence Alignment in Text Simplification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7943–7960.

Tomoyuki Kajiwara, Biwa Miura, and Yuki Arase. 2020. Monolingual Transfer Learning via Bilingual Translators for Style-Sensitive Paraphrase Generation. In *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence*, pages 8042–8049.

Akihiro Katsuta and Kazuhide Yamamoto. 2018. Crowdsourced Corpus of Sentence Simplification with Core Vocabulary. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, pages 461–466.

Taku Kudo and John Richardson. 2018. SentencePiece: A Simple and Language Independent Subword Tokenizer and Detokenizer for Neural Text Processing. In *Proceedings of the 2018 Conference* *on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 66–71.

Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2019. BioBERT: A Pre-trained Biomedical Language Representation Model for Biomedical Text Mining. *Bioinformatics*, 36(4):1234–1240.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled Weight Decay Regularization. In *Proceedings of the 2018 International Conference on Learning Representations*.

Junyu Luo, Junxian Lin, Chi Lin, Cao Xiao, Xinning Gui, and Fenglong Ma. 2022. Benchmarking Automated Clinical Language Simplification: Dataset, Algorithm, and Evaluation. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 3550–3562.

Mounica Maddela and Wei Xu. 2018. A Word-Complexity Lexicon and A Neural Readability Ranking Model for Lexical Simplification. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3749–3760.

Louis Martin, Angela Fan, Éric de la Clergerie, Antoine Bordes, and Benoît Sagot. 2022. MUSS: Multilingual Unsupervised Sentence Simplification by Mining Paraphrases. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1651–1664.

Takumi Maruyama and Kazuhide Yamamoto. 2018. Simplified Corpus with Core Vocabulary. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, pages 1153–1160.

Daiki Nishihara and Tomoyuki Kajiwara. 2020. Word Complexity Estimation for Japanese Lexical Simplification. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 3114–3120.

Chunguang Pan, Bingyan Song, Shengguang Wang, and Zhipeng Luo. 2021. DeepBlueAI at SemEval-2021 Task 1: Lexical Complexity Prediction with A Deep Ensemble Approach. In *Proceedings of the 15th International Workshop on Semantic Evaluation*, pages 578–584.

Tarek Sakakini, Jong Yoon Lee, Aditya Duri, Renato F.L. Azevedo, Victor Sadauskas, Kuangxiao Gu, Suma Bhat, Dan Morrow, James Graumlich, Saqib Walayat, Mark Hasegawa-Johnson, Thomas Huang, Ann Willemsen-Dunlap, and Donald Halpin. 2020. Context-Aware Automatic Text Simplification of Health Materials in Low-Resource Domains. In *Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis*, pages 115–126.

Renliang Sun and Xiaojun Wan. 2022. SimpleBERT: A Pre-trained Model That Learns to Generate Simple Words. *arXiv:2204.07779*.

Renliang Sun, Wei Xu, and Xiaojun Wan. 2023. Teaching the Pre-trained Model to Generate Simple Texts for Text Simplification. In *Proceedings of the 2023 Findings of the Association for Computational Linguistics*, pages 9345–9355.

Arseny Tolmachev, Daisuke Kawahara, and Sadao Kurohashi. 2018. Juman++: A Morphological Analysis Toolkit for Scriptio Continua. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 54–59.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. In *Proceedings of the 31st Conference on Neural Information Processing Systems*, pages 5998–6008.

Wei Xu, Chris Callison-Burch, and Courtney Napoles. 2015. Problems in Current Text Simplification Research: New Data Can Help. *Transactions of the Association for Computational Linguistics*, 3:283–297.

Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch. 2016. Optimizing Statistical Machine Translation for Text Simplification. *Transactions of the Association for Computational Linguistics*, 4:401–415.

Tatsuya Zetsu, Tomoyuki Kajiwara, and Yuki Arase. 2022. Lexically Constrained Decoding with Edit Operation Prediction for Controllable Text Simplification. In *Proceedings of the 2022 Workshop on Text Simplification, Accessibility, and Readability*, pages 147–153.

Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter J. Liu. 2020. PEGASUS: Pre-Training with Extracted Gap-Sentences for Abstractive Summarization. In *Proceedings of the 37th International Conference on Machine Learning*.